

Oracles and query lower bounds in generalised probabilistic theories

Ciarán M. Lee^{‡,1}, John H. Selby^{*,•,2} and Howard Barnum^{†,§}

[‡] *Department of Physics and Astronomy, University College London, UK.*

^{*} *University of Oxford, Department of Computer Science, OX1 3QD, UK.*

[•] *Imperial College London, London SW7 2AZ, UK.*

[†] *QMATH, Department of Mathematical Sciences, University of Copenhagen, Denmark.*

[§] *Department of Physics and Astronomy, University of New Mexico, Albuquerque, USA.*

We investigate the connection between interference and computational power within the operationally defined framework of generalised probabilistic theories. To compare the computational abilities of different theories within this framework we show that any theory satisfying three natural physical principles possess a well-defined oracle model. Indeed, we prove a subroutine theorem for oracles in such theories which is a necessary condition for the oracle to be well-defined. The three principles are: causality (roughly, no signalling from the future), purification (each mixed state arises as the marginal of a pure state of a larger system), and strong symmetry (existence of non-trivial reversible transformations). Sorkin has defined a hierarchy of conceivable interference behaviours, where the order in the hierarchy corresponds to the number of paths that have an irreducible interaction in a multi-slit experiment. Given our oracle model, we show that if a classical computer requires at least n queries to solve a learning problem, then the corresponding lower bound in theories lying at the k th level of Sorkin's hierarchy is $\lceil n/k \rceil$. Hence, lower bounds on the number of queries to a quantum oracle needed to solve certain problems are not optimal in the space of all generalised probabilistic theories, although it is not yet known whether the optimal bounds are achievable in general. Hence searches for higher-order interference are not only foundationally motivated, but constitute a search for a computational resource beyond that offered by quantum computation.

Landauer's Principle [1] states that any logically irreversible processing or manipulation of information, such as the erasure of a bit, must always be accompanied by an entropy increase in the environment of the system processing the information. That is, information is intimately tied to the physical system it embodies and is hence bound by physical law—alternatively, *information is physical*. If information processing—or computation—is bound by physical law, then the ultimate limits of computation should be derivable from natural physical principles. Indeed, the advent of quantum computation demonstrated that different physical principles convey different limits on computational power. This naturally leads to the question of what general relationships hold between computational power and physical principles. This question has recently been studied in the framework of generalised probabilistic theories [2, 3, 4, 5, 6], which contains operationally-defined physical theories that generalise the probabilistic formalism of quantum theory. By studying how computational power varies as the underlying physical theory is changed, one can determine the connection between physical principles and computational power in a manner not tied to the specific mathematical manifestation of a particular principle within a theory.

All previous research into computation within the generalised probabilistic theory framework has thus far focused on deriving general bounds on computational ability from natural

¹ Electronic address: ciaran.lee@ucl.ac.uk

² Electronic address: john.selby08@imperial.ac.uk

physical principles³. No work to date has tied a computational advantage directly to a physical principle. For instance, it is known that quantum interference between computational paths is a resource for post-classical computation [7], but it is not clear whether the presence of interference in a general theory entails post-classical computation [8], or whether post-quantum interference behaviour is in general a resource for post-quantum computation [4]. The former point concerns whether it is just the particular mathematical description of interference in Hilbert space which can be exploited to provide an advantage over classical computation or whether such a statement can be seen to directly follow from the observation of interference in nature, and the latter concerns whether “more” interference implies “more” computational power.

Indeed, as first noted by Rafael Sorkin [9, 10], there is a limit to quantum interference—at most pairs of computational paths can ever interact in a fundamental way. Sorkin has defined a hierarchy of operationally conceivable interference behaviours—currently under experimental investigation [11, 12, 13, 14]—where classical theory is at the first level of the hierarchy and quantum theory belongs to the second. Informally, the order in the hierarchy corresponds to the number of paths that have an irreducible interaction in a multi-slit experiment. Given this definition, one can investigate the role interference plays in computation in a theory-independent manner by asking whether theories at level k possess a computational advantage over theories at level $k - 1, k - 2, \dots$.

One usually demonstrates the existence of a quantum advantage over classical computation using *oracles*. Indeed, the Deutsch-Jozsa problem [15] is such an example: given a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$, one is asked to determine whether it is constant (same output for all inputs) or balanced (0 on exactly half of the inputs)—promised that it is one of these cases. Performance is quantified by the number of queries to an oracle, which implements f on any desired input, needed to solve the problem. A classical computer requires $2^{n-1} + 1$ queries, but Deutsch and Jozsa showed that a quantum computer can solve the problem in a single query. Hence, to compare the computational abilities of different theories within the framework a well-defined oracle model is needed. We show that such a model can be defined in any theory satisfying the following physical principles: *causality* (roughly, no signalling from the future), *purification* (each mixed state arises as the marginal of a pure state of a larger system), and *strong symmetry* (existence of non-trivial reversible transformations), additionally, we demand *computationally consistent composition* (the act of bringing systems together cannot solve a hard problem). Moreover, we prove a subroutine theorem for theories satisfying these principles. That is, we show that having access to an oracle for a particular decision problem which can be efficiently solved in a given theory does not provide any more computational power than just using the efficient algorithm itself. Such a result was proved for quantum theory by Bennett et al. in [16], and is a necessary condition for a well-defined oracle model.

Given this oracle model, we investigate whether lower bounds on the number of queries needed by a quantum computer to solve certain computational problems can be reduced in theories which possess higher-order interference and satisfy the principles discussed above. Indeed, we generalise results due to Meyer and Pommersheim [17] and show that if a lower bound to a specific query problem is n using a classical computer, then the corresponding lower bound in theories with k th order interference is $\lceil n/k \rceil$. As quantum theory only exhibits second order interference, theories with post-quantum interference allow for post-quantum computation. For example, in the specific generalisation of Deutsch’s problem where we are asked to determine the parity of a function $f : \{1, \dots, k\} \rightarrow \{0, 1\}$, $\lceil k/2 \rceil$ quantum queries are needed, but we show that in any theory satisfying our principles which has k th order

³With the exception of [5], which constructs a theory capable of post-quantum computation. However, whether this computational advantage is directly tied to a simple physical principle remains unclear.

interference it may be the case that the parity can be determined with a single query. That is, lower bounds for certain query problems in quantum computation can be improved by modifying interference behaviour in an operationally conceivable manner. Hence searches for higher-order interference are not only foundationally motivated, but constitute a search for a computational resource beyond that offered by quantum computation. An important direction for future work is to determine whether in theories satisfying these principles it is possible to find an algorithm that reaches this lower bound. We discuss potential ways to do this in the conclusion.

Other authors have considered computation beyond the usual quantum formalism from a different perspective to the one employed here. For example, Aaronson has considered alternate modifications of quantum theory, such as a hidden variable model in which the history of hidden states can be read out by the observer, and—together with collaborators in [18]—a model in which one is given the ability to perform certain unphysical non-collapsing measurements. Both of these models have been shown to entail computational speed-ups over the usual quantum formalism. Additionally, Bao et al. [19] have investigated computation in modifications of quantum theory suggested by the black hole information loss paradox and have shown the ability to signal faster than light in such theories is intimately linked to a speed-up over standard quantum theory in searching an unstructured database. In contrast, the generalised probabilistic theory framework employed here allowed for an investigation of query lower bounds and computational advantages in alternate theories that are physically reasonable and which, for instance, do not allow for superluminal signalling [20], cloning [21], or other ad-hoc assumptions.

The paper is organised as follows. In section 1, the generalised probabilistic theory framework will be introduced along with our physical principles and the definition of higher-order interference. In section 2 the oracle model will be introduced and defined and the subroutine theorem will be stated, with the proof presented in appendix C. Finally, in section 3 we derive lower bounds on query problems that directly follow from our principles.

1 The framework

One of the fundamental requirements of a physical theory is that it should provide a consistent account of experimental observations. This viewpoint underlies the framework of generalised probabilistic theories [20, 22, 23, 24]. A generalised probabilistic theory specifies a set of laboratory devices that can be connected together in different ways to form experiments and specifies probability distributions over experimental outcomes. A device comes equipped with input ports, output ports, and a classical pointer. When a device is used in an experiment, the classical pointer comes to rest in one of a number of positions, indicating an outcome has occurred. Intuitively, one envisages *physical systems* passing between the ports of these devices. Such physical systems come in different types, denoted A, B, \dots . One constructs experiments by composing devices both sequentially and in parallel, and when composed sequentially, types must match.

In this framework, closed circuits—those with no unconnected ports and no cycles—define probabilities. Devices that yield the same probabilities in all closed circuits are identified. The set of equivalence classes of devices with no input ports are called *states*, no output ports *effects* and both input and output ports *transformations*. The set of all states of system A is denoted $\text{St}(A)$, the set of all effects on B is denoted $\text{Eff}(B)$ and the set of transformations between systems A and B is denoted $\text{Transf}(A, B)$. Using standard operational arguments [2, 22, 20], one can show that the set of states, effects and transformations each give rise to a real vector space with transformations and effects acting linearly on the real vector space

of states. We assume in this work that all vector spaces are finite dimensional⁴.

A state is said to be *pure* if it does not arise as a *coarse-graining* of other states⁵; a pure state is one for which we have maximal information. A state is *mixed* if it is not pure. A state is *maximally mixed* if every state appears in its coarse graining. Similarly, one says a transformation is pure if it does not arise as a coarse-graining of other transformations. It can be shown that reversible transformations preserve pure states [25].

The following ‘Dirac-like’ notation ${}_A|s_i\rangle$ will be used to represent a state⁶ of system A , and $\langle e_r|_B$ to represent an effect on B . Here i and r represent the position of the classical pointer associated to the device the prepares the state and performs the measurement, respectively. The full measurement is defined by the collection $\{\langle e_r|\}_r$. States, effects, and transformations can be represented diagrammatically:

$$\left(\begin{array}{c} \text{---} s_i \end{array} \right) \begin{array}{c} A \\ \text{---} T \end{array} \begin{array}{c} B \\ \text{---} e_r \end{array} \left(\right) := \langle e_r|_B T_A |s_i\rangle$$

The above diagram represents the joint probability of preparing state $|s_i\rangle$, acting with transformation T and registering outcome r for the measurement $\{\langle e_r|\}_r$. In the above, the wires represent physical systems, with their type denoted by the letter above them. This diagrammatic representation was inspired by categorical quantum mechanics [26, 27].

In the rest of the paper, it will be assumed that all theories satisfy the following physical principles.

Definition 1.0.1 (Causality [22]). *A theory is said to be causal if there exists a unique deterministic effect $\langle \mathbb{I} |$ for every system, such that $\sum_r \langle e_r | = \langle \mathbb{I} |$ for all measurements, $\{\langle e_r | \}_r$.*

Mathematically, the principle of causality is equivalent to the statement: “Probabilities of present experiments are independent of future measurement choices”. In causal theories, all states are *normalised* [22]. That is, $\langle \mathbb{I} | s \rangle = 1$ for all $|s\rangle$. Moreover, the unique deterministic effect allows one to define a notion of *marginalisation* for multi-partite states.

Definition 1.0.2 (Purification [22]). *Given a state ${}_A|s\rangle$ there exists a system B and a pure state ${}_{AB}|\psi\rangle$ on AB such that ${}_A|s\rangle$ is the marginalisation of ${}_{AB}|\psi\rangle$:*

$$\left(\begin{array}{c} A \\ \text{---} \psi \\ B \\ \text{---} \mathbb{I} \end{array} \right) = \left(\begin{array}{c} A \\ \text{---} s \end{array} \right)$$

Moreover, the purification ${}_{AB}|\psi\rangle$ is unique up to reversible transformations on the purifying system, B . That is if two states $|\psi\rangle_{AB}$ and $|\psi'\rangle_{AB}$ purify $|s\rangle_A$, then there exists a reversible transformation T_B on system B such that $|\psi\rangle_{AB} = (\mathbb{I}_A \otimes T_B)|\psi'\rangle_{AB}$.

As pure states are those about which we have maximal information, the purification principle formalises the statement that each state of incomplete information arises in an essentially unique way due to a lack of information about a larger system. Roughly, purification can be thought of as a statement of “information conservation” [28]; any missing information about the state of a given system can always be traced back to lack of information of some environment system.

We introduce one final principle which ensures that the computational power of a theory is compatible with composition, that is, we demand that the mere act of bringing systems

⁴Operationally this can be seen as saying that one does not need to perform an infinite number of distinct experiments to characterise states

⁵The process $\{\mathcal{U}_j\}_{j \in Y}$, where j index the positions of the classical pointer, is a coarse-graining of the process $\{\mathcal{E}_i\}_{i \in X}$ if there is a disjoint partition $\{X_j\}_{j \in Y}$ of X such that $\mathcal{U}_j = \sum_{i \in X_j} \mathcal{E}_i$.

⁶or, more accurately, the real vector corresponding to the state.

together should not solve a hard problem that could not be solved by using the systems independently [2]. If one could “lose” or “gain” information when physical systems are brought together, then one could potentially use this new global degree of freedom, representing the increase or decrease of information, to hide solutions to a hard computational problem. We formalise this as follows:

Definition 1.0.3 (Computationally consistent composition). *This consists of two constraints on parallel composition: i) the product of pure states is pure, ii) the product of maximally mixed states is maximally mixed.*

The first of these formalises the intuitive idea that if one has maximal information about two systems, then one should not “lose” information by bringing them together. The existence of a maximally mixed state, that is, a state about which we have minimal information, is guaranteed for each system by purification [22]. The second constraint guarantees that if we have no information about two individual systems then we have no information about the composite.

The purification principle, in conjunction with causality and the constraints on parallel composition discussed above, implies many quantum information processing [22] and computational primitives [4]. Examples include teleportation, no information without disturbance, and no-bit commitment [22]. Moreover, purification also leads to a well-defined notion of thermodynamics [25, 29, 30]. Quantum theory—both on complex and real Hilbert spaces—satisfies purification as does Spekkens’ toy model [31, 32] fermionic quantum theory [33, 34], a superselected version of quantum theory known as double quantum theory [30], and a recent extension of classical theory to the theory of coherent d -level systems, or codits [25].

Pure states $\{|s_i\rangle\}_{i=1}^n$ are called *perfectly distinguishable* if there exists a measurement, corresponding to effects $\{e_j\}_{j=1}^n$, with the property that $(e_j|s_i) = \delta_{ij}$ for all i, j .

Definition 1.0.4 (Strong symmetry [35, 36]). *A theory satisfies strong symmetry if, for any two n -tuples of pure and perfectly distinguishable states $\{|\rho_i\rangle\}, \{|\sigma_i\rangle\}$, there exists a reversible transformation T such that $T|\rho_i\rangle = |\sigma_i\rangle$ for $i = 1, \dots, n$.*

The following consequences of the above principles, proved in [8], will be required to define oracles in section 2.

Definition 1.0.5. *Given a set of pure and perfectly distinguishable states $\{|i\rangle\}$, and a set of transformations $\{T_i\}$, define a controlled transformation $C\{T_i\}$ as:*

$$\begin{array}{c} \text{⌞} i \\ \text{⌞} \sigma \end{array} \begin{array}{c} \boxed{C} \\ \boxed{\{T_i\}} \end{array} \begin{array}{c} \text{---} \\ \text{---} \end{array} = \begin{array}{c} \text{⌞} i \\ \text{⌞} \sigma \end{array} \begin{array}{c} \text{---} \\ \boxed{T_i} \end{array} \begin{array}{c} \text{---} \\ \text{---} \end{array} \quad \forall i, |\sigma\rangle \quad (1.0.1)$$

The top system and lower systems are referred to as the control and target respectively.

Note that classical controlled transformations—where the control is measured and, conditioned on the outcome, a transformation is applied to the target—exist in any causal theory with sufficient distinguishable states [22]. However, such transformations are in general not reversible [8].

Theorem 1.0.6 ([8] Theorem 2). *In any theory satisfying i) causality, ii) purification, iii) strong symmetry, iv) product of pure states is pure, and in which there exists a set of n pure and perfectly distinguishable states, there exists a reversible controlled transformation for any collection of n reversible transformations $\{T_i\}_{i=1}^n$.*

Every controlled unitary transformation in quantum theory has a *phase kick-back* mechanism [15, 37]. Such mechanisms form a vital component of most quantum algorithms [37]. It was shown in [8] that a *generalised* phase kick-back mechanism exists in any theory satisfying the above physical principles.

Theorem 1.0.7 ([8] Lemma 2). *Given an $|s\rangle$ such that $T_i|s\rangle = |s\rangle$, $\forall i$, there exists a reversible transformation Q_s such that*

$$\begin{array}{c} \sigma \\ \hline \end{array} \begin{array}{c} \boxed{C} \\ \hline \end{array} \begin{array}{c} \hline \\ s \end{array} = \begin{array}{c} \sigma \\ \hline \end{array} \begin{array}{c} \boxed{Q_s} \\ \hline \end{array} \begin{array}{c} \hline \\ s \end{array} \quad \forall |\sigma\rangle \quad (1.0.2)$$

where Q_s is phase transformation:

$$\begin{array}{c} \boxed{Q_s} \\ \hline \end{array} \begin{array}{c} \hline \\ i \end{array} = \begin{array}{c} \hline \\ i \end{array} \quad \forall i$$

Moreover, every phase transformation can arise via a generalised phase kick-back mechanism

1.1 Post-quantum interference

In the sections that follow, we will connect post-quantum, or higher-order, interference to post-quantum computation; investigating whether “more” interference implies “more” computational power. The definition of higher-order interference that we present here takes its motivation from the set-up of multi-slit interference experiments. In such experiments a particle (a photon or electron, say) passes through slits in a physical barrier. By blocking some of the slits and repeating the experiment many times, one can build up an interference pattern on a screen placed behind the physical barrier. Informally, a theory has “ n th order interference” if one can generate interference patterns in an n -slit experiment which cannot be created in any experiment with only m -slits, for all $m < n$.

More precisely, this means that the interference pattern created on the screen cannot be written as a particular linear combination of the interference patterns generated when different subsets of slits are open and closed. In the standard two slit experiment, quantum interference corresponds to the statement that the interference pattern can’t be written as the sum of single slit patterns:

$$\begin{array}{c} | \\ | \\ | \end{array} \neq \begin{array}{c} | \\ | \\ | \end{array} + \begin{array}{c} | \\ | \\ | \end{array}$$

It was first shown by Sorkin [9, 10] that—at least for ideal experiments [38]—quantum theory is limited to the $n = 2$ case. That is, the interference pattern created in a three—or more—slit experiment *can* be written in terms of the two and one slit interference patterns obtained by blocking some of the slits. Schematically:

$$\begin{array}{c} | \\ | \\ | \end{array} = \begin{array}{c} | \\ | \\ | \end{array} + \begin{array}{c} | \\ | \\ | \end{array} + \begin{array}{c} | \\ | \\ | \end{array} - \begin{array}{c} | \\ | \\ | \end{array} - \begin{array}{c} | \\ | \\ | \end{array} - \begin{array}{c} | \\ | \\ | \end{array}$$

The minus signs in the above account for over-counting of the open slits. If a theory does not have n th order interference then one can show it will not have m th order interference, for any $m > n$ [9]. As such, one can classify theories according to their maximal order of interference, k . For example quantum theory lies at $k = 2$ and classical theory at $k = 1$.

Consider the state of the particle just before it passes through the slits. For every slit, there should exist states such that the particle is definitely found at that slit, if one were to measure it. Mathematically, this means that there exists a face⁷ [35] of the state space, such that all states in this face give unit probability for the “yes” outcome of the two outcome “is the particle at this slit?” measurement. These faces will be labelled F_i , one for each of the n slits $i \in \{1, \dots, n\}$. As the slits should be perfectly distinguishable, the faces associated to each slit should be orthogonal. This can be achieved by letting the slits be in one-to-one correspondence with a set of pure and perfectly distinguishable states.

One can additionally ask coarse grained questions of the form “Is the particle found among a certain subset of slits, rather than somewhere else?”. The set of states that give outcome “yes” with probability one must contain all the faces associated with each slit in the subset. Hence the face associated to the subset of slits $I \subseteq \{1, \dots, n\}$ is the smallest face containing each face in this subset $F_I := \bigvee_{i \in I} F_i$. That is, F_I is the face generated by the pure and perfectly distinguishable states identified by the subset I . The face F_I contains all those states which can be found among the I slits. The experiment is “complete” if all states in the state space (of a given type A) can be found among some subset of slits. That is, if $F_{12\dots n} = \text{St}(A)$.

Higher-order interference was initially formalised by Rafael Sorkin in the framework of Quantum Measure Theory [9] but has more recently been adapted to the setting of generalised probabilistic theories in [35, 8, 39, 40]. A straightforward translation to this setting describes the order of interference in terms of probability distributions corresponding to interference patterns generated in the different experimental setups (which slits are open, etc.) [8, 40]. However, given the principles imposed in the previous section, it is possible to define physical transformations that correspond to the action of opening and closing certain subsets of slits. In this case, there is a more convenient (and equivalent, given our principles) definition in terms of such transformations [35].

Given N slits, labelled $1, \dots, N$, these transformations will be denoted P_I , where $I \subseteq \{1, \dots, N\}$ corresponds to the subset of slits which are not closed. In general one expects that $P_I P_J = P_{I \cap J}$, as only those slits belonging to both I and J will not be closed by either P_I or P_J . This intuition suggests that these transformations should correspond to projectors (i.e. idempotent transformations $P_I P_I = P_I$). Given the principles imposed in this paper, this is indeed the case.

Theorem 1.1.1. *In any theory satisfying the principles introduced in the previous section, the projector onto a face generated by a subset of pure and perfectly distinguishable states is an allowed transformation in the theory.*

The proof of theorem 1.1.1 is presented in appendix A. Given this structure, one can define the maximal order of interference as follows [35, 4].

Definition 1.1.2. *A theory satisfying the principles imposed in this section has maximal order of interference k if, for any $N \geq k$, one has:*

$$\mathbb{1}_N = \sum_{\substack{I \subseteq N \\ |I| \leq k}} \mathcal{C}(k, |I|, N) P_I$$

where $\mathbb{1}_N$ is the identity on a system with N pure and perfectly distinguishable states and

$$\mathcal{C}(k, |I|, N) := (-1)^{k-|I|} \binom{N-|I|-1}{k-|I|}$$

⁷A face is a convex set with the property that if $px + (1-p)y$, for some $p \in (0, 1)$, is an element then x and y are also both elements.

The factor $\mathcal{C}(k, |I|, N)$ in the above definition corrects for the overlaps that occur when different combinations of slits are open and closed. For $k = N$, the above reduces to the expected expression $\mathbb{1}_h = P_{\{1, \dots, k\}}$, that is, the identity is given by the projector with all slits open. The case $N = k + 1$ corresponds to $\mathcal{C}(k, |I|, k + 1) = (-1)^{k - |I|}$, corresponding to the situation depicted in the above diagrams, as well as the one most commonly discussed in the literature [9, 39].

Instead of working directly with these physical projectors, it is mathematically convenient to work with the (generally) unphysical transformations corresponding to projecting onto the “coherences” of a state. Consider the example of a qutrit in quantum theory, the projector $P_{\{0,1\}}$ projects onto a two dimensional subspace:

$$P_{\{0,1\}} :: \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} \\ \rho_{10} & \rho_{11} & \rho_{12} \\ \rho_{20} & \rho_{21} & \rho_{22} \end{pmatrix} \mapsto \begin{pmatrix} \rho_{00} & \rho_{01} & 0 \\ \rho_{10} & \rho_{11} & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

whilst the coherence-projector $\omega_{\{0,1\}}$ projects only onto the coherences in that two dimensional subspace:

$$\omega_{\{0,1\}} :: \begin{pmatrix} \rho_{00} & \rho_{01} & \rho_{02} \\ \rho_{10} & \rho_{11} & \rho_{12} \\ \rho_{20} & \rho_{21} & \rho_{22} \end{pmatrix} \mapsto \begin{pmatrix} 0 & \rho_{01} & 0 \\ \rho_{10} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

That is, $\omega_{\{0,1\}}$ corresponds to the linear combination of projectors: $P_{\{0,1\}} - P_{\{0\}} - P_{\{1\}}$.

There is a coherence-projector ω_I for each subset of slits $I \subseteq \{1, \dots, N\}$, defined in terms of the physical projectors:

$$\omega_I := \sum_{\tilde{I} \subseteq I} (-1)^{|I| + |\tilde{I}|} P_{\tilde{I}}.$$

These have the following useful properties, which were proved in [4, 35].

Lemma 1.1.3. *An equivalent definition of the maximal order of interference, k , is:*

$$\mathbb{1}_N = \sum_{I, |I|=1}^k \omega_I, \text{ for all } N \geq k.$$

The above lemma implies that any state (indeed, any vector in the vector space generated by the states) in a theory with maximal order of interference k can be decomposed in a form reminiscent of a rank k tensor:

$$|s\rangle = \sum_{I, |I|=1}^k \omega_I |s\rangle = \sum_{I, |I|=1}^k |s_I\rangle. \quad (1.1.1)$$

This decomposition can be thought of as a generalised superposition, as it manifestly describes the coherences between different subsets of perfectly distinguishable states (the analogue of a basis in quantum theory) present in a given state. This will be very useful when discussing oracle queries in the following section.

2 Oracles

In classical computation, an *oracle* is a total function $f : \mathbb{N} \rightarrow \{0, 1\}$. The x is said to be in an oracle O if $f(x) = 1$, hence oracles can decide membership in a language. In quantum computation oracle queries are represented by a family $\{G_n\}$ of quantum gates, one for each

query length. Each G_n is a unitary transformation acting on $n + 1$ qubits, whose effect on the computational basis is in general given by

$$G_n|x, a\rangle = |x, a \oplus f_n(x)\rangle$$

for all $x \in \{0, 1\}^n$ and $a \in \{0, 1\}$, where f_n is some Boolean function that represents the specific oracle under consideration. One can think of a quantum oracle as a controlled unitary transformation which, when queried by state $|x\rangle$ in the control register, applies a unitary—chosen from a set of two unitaries according to the value $f_n(x)$ —to the target register. A specific example of a quantum oracle is the following controlled unitary:

$$U_f = |0\rangle\langle 0| \otimes Z^{f(0)} + |1\rangle\langle 1| \otimes Z^{f(1)}, \quad (2.0.1)$$

with Z a Pauli matrix, $f : \{0, 1\} \rightarrow \{0, 1\}$ a function encoding some decision problem and $Z^0 := \mathbb{I}$.

As was briefly mentioned in [8], the results of theorem 1.0.6 provide a way to define computational oracles in any theory satisfying our assumptions. An oracle in such theories corresponds to a reversible controlled transformation⁸ where the set of transformations $\{T_{i,f(i)}\}$ being controlled depend on a function $f : \{1, \dots, n\} \rightarrow \{0, 1\}$ encoding a decision problem of interest.

One can schematically represent the problems that can be solved by a specific computational model with access to an oracle using the language of complexity classes. Let \mathbf{C} and \mathbf{B} be complexity classes, then $\mathbf{C}^{\mathbf{B}}$ denotes the class \mathbf{C} with an oracle for \mathbf{B} (see [41] for formal definitions). We can think of $\mathbf{C}^{\mathbf{B}}$ as the class of languages decided by a computation which is subject to the restrictions and acceptance criteria of \mathbf{C} , but allowing an extra kind of computational step: an oracle for any desired language $\mathcal{L} \in \mathbf{B}$ that may be queried during the computation, where each query counts as a single computational step.

A natural question is whether having access to an oracle⁹ for a particular decision problem which can be efficiently solved in a given theory provide any more computational power than just using the efficient algorithm? If we schematically denote the class of problems efficiently solvable by a particular theory \mathbf{G} by¹⁰ \mathbf{BGP} , this questions can be phrased as: “is \mathbf{BGP} is closed under *subroutines*”? Here \mathbf{BGP} is the analogue of the well-known class of problems efficiently solvable by a quantum computer, \mathbf{BQP} . Another way to pose this question is to ask whether $\mathbf{BGP}^{\mathbf{BGP}} = \mathbf{BGP}$ for \mathbf{G} satisfying our principles.

There exist complexity classes for which this is not the case, for example, \mathbf{NP} ¹¹. But, intuitively, one would expect it to hold in a sensible physical theory where computation is performed with circuits. A potential issue arises when one compares the performance of the oracle implementation to that of the efficient algorithm when both are used as subroutines in another computational algorithm¹². As we have seen above, oracles can be queried in superposition, but one does not usually query an algorithm for a particular decision problem in superposition; one merely prepares the state corresponding to a particular bit string and uses the algorithm to determine whether or not that bit string is in the language in question. For simplicity, we say the efficient algorithm accepts an input if a measurement of the first outcome system yields outcome $|0\rangle$ with probability¹³ greater than $2/3$. This is the same acceptance condition imposed in quantum computation.

⁸There could be many distinct transformations that have the same behaviour on a set of control states. As long as one fixes which transformation corresponds to the oracle, this is not a problem.

⁹Represented as some function which can be queried using said oracle

¹⁰See references [2, 3, 5] for a rigorous definition of this class.

¹¹If one assumes that the polynomial hierarchy doesn’t collapse.

¹²Here, an algorithm consists of a poly-size uniform circuit. See [2] for the formal definitions.

¹³This can be amplified to $1 - 2^{-q}$, where q a polynomial in the size of the circuit, by running the circuit in parallel a polynomial number of times. Again, see [2].

We therefore need to know whether every **BGP** algorithm for a decision problem admits a subroutine having the characteristics of an oracle for that decision problem. Such a result was proved in the quantum case by Bennett et al. in [16]. The following theorem shows that it is also true for theories satisfying our principles.

Theorem 2.0.1. *Consider a theory \mathbf{G} which satisfies the principles outlined in section 1. Given an algorithm $\{A_{|x|}\}$ for a decision problem in **BGP**, one can always construct a circuit family $\{G_{|x|}\}$, consisting of reversible transformations from \mathbf{G} , which, with high probability, functions as an oracle for that particular decision problem. Schematically, we have $\mathbf{BGP}^{\mathbf{BGP}} = \mathbf{BGP}$.*

Proof. See Appendix C □

In quantum theory there is an equivalent view of oracles in terms of phase transformations, this can be seen as a result of the phase kick-back algorithm [37, 15]. In the quantum example above, the phase kick-back for U_f amounts to first rewriting U_f as:

$$U_f = \mathbb{I} \otimes |0\rangle\langle 0| + Z^{f(0) \oplus f(1)} \otimes |1\rangle\langle 1|. \quad (2.0.2)$$

Inputting $|1\rangle$ into the second qubit results in a ‘kicked-back’ phase of $Z^{f(0) \oplus f(1)}$ on the first qubit. The value of $f(0) \oplus f(1)$ can then be measured by preparing the first qubit in the state $|+\rangle$ and then measuring it in the $\{|+\rangle, |-\rangle\}$ basis. Therefore providing the value $f(0) \oplus f(1)$ in a single query of the oracle—a feat impossible on a classical computer [17].

In our more general setting, an analogue of the above holds via Theorem 1.0.7. That is, in theories satisfying our assumptions, as the transformations $T_{i,f(i)}$ depend on the value of $f(i)$, so too can the controlled transformation and the kicked-back phase. That is, in theories with non-trivial phases, i.e. non-classical theories, the phase kick-back of an oracle can encode information about the value $f(i)$ for all i . Hence—as in the quantum example above—from the point of view of querying the oracle, one can reduce all considerations involving the controlled transformation to those involving the kicked-back phase, which shall be denoted \mathcal{O}_f .

As was shown in section 1, all states in theories satisfying our principles can be decomposed as $|s\rangle = \sum_I |s_I\rangle$, with $I \subseteq \{1, \dots, n\}$, where $\{1, \dots, n\}$ labels the set of pure and perfectly distinguishable states defining the action of a give oracle. Hence, oracles can not only be queried using a set of pure and perfectly distinguishable states, but also using generalised superposition states—those with non-trivial coherences between different subsets of slits. In fact, the quantum speed-up in the above example came precisely from the fact that one can query in superposition, hence extracting the value $f(0) \oplus f(1)$ in a single query. To ensure that answers to hard to solve problems are not smuggled into the definition of oracles in generalised theories, we must put conditions on which phase transformations correspond to ‘reasonable’ oracles.

Definition 2.0.2 (Oracle). *An oracle for a function f is defined as a phase transformation \mathcal{O}_f with the additional constraint that given functions f and g such that $f(i) = g(i)$ for all $i \in I$, the oracles corresponding to them must satisfy*

$$\mathcal{O}_f |s_I\rangle = \mathcal{O}_g |s_I\rangle$$

for all $|s_I\rangle = \omega_I |s\rangle$.

This ensures one cannot learn about the value $f(j)$ when querying using a state with no probability of being found in $|j\rangle$. Given this definition of an oracle we can consider how their computational power depends on the order of interference of the theory.

3 Lower bounds from useless queries

In this section we generalise results of Ref. [17] where Meyer and Pommersheim derived a relation between quantum and classical query complexity lower bounds by introducing the concept of a “useless” quantum query to the setting of GPTs satisfying our principles. They considered *learning problems* in which one is given an element from a class of functions with the same domain and range, chosen with some arbitrary—but known—prior distribution, where the task is to determine to which specific subclass the chosen function belongs.

More formally we can define a learning problem as follows:

Definition 3.0.1 (Learning problems). *Given a set of functions $\mathcal{C} \subseteq \{0,1\}^X$ where X is some finite set¹⁴, and, a partitioning of \mathcal{C} into disjoint subsets $\mathcal{C} = \bigsqcup_{j \in J} \mathcal{C}_j$ labeled by $j \in J$. The aim of the learning problem is to determine which partition \mathcal{C}_j a particular function $f \in \mathcal{C}$ belongs to, given a prior μ with which the function is chosen from among the \mathcal{C}_j ’s. A particular learning problem is therefore defined by the triple, $(\mathcal{C}, \{\mathcal{C}_j\}, \mu)$.*

One can only access information about the function by querying an oracle, which, when presented with an element from the domain, outputs the corresponding element of the range assigned by the chosen function. Meyer and Pommersheim showed that if n queries to a classical oracle reveal no information about which function was chosen¹⁵ then neither do $n/2$ queries to a quantum oracle. Hence $\lceil n/2 \rceil + 1$ quantum queries constitute a lower bound.

Many important query problems are examples of learning problems. For instance, **PARITY**, a generalisation of Deutsch’s problem [15] which asks for the parity of a function $f : \{1, \dots, N\} \rightarrow \{0,1\}$ can be written as a learning problem. Indeed, partition the class of all such functions into two subclasses, one in which all functions have parity 0 and the other 1, and choose the function with a prior probability of $1/2$. In this case, $N - 1$ classical queries do not provide any information about the parity, hence at least $\lceil (N - 1)/2 \rceil + 1$ quantum queries are needed to solve the problem. Indeed, $\lceil (N - 1)/2 \rceil + 1$ quantum queries are sufficient¹⁶.

In this section we generalise Meyer and Pommersheim’s result to the case of oracle queries in the generalised probabilistic theory framework presented in the previous section. We prove that if n queries to a classical oracle reveal no information about which function was chosen then neither do n/k queries in a generalised theory satisfying the principles introduced in section 1 and which has maximal order of interference k . Hence a lower bound to determining the function is $\lceil n/k \rceil + 1$ queries in theories with k th order interference. Hence, in the specific generalisation of Deutsch’s problem where we are asked to determine the parity of a function $f : \{1, \dots, k\} \rightarrow \{0,1\}$, $\lceil k/2 \rceil$ quantum queries are needed, but in a theory with k th order interference it may be the case that the parity can be determined with a single query.

We can now formally define what it means for n classical queries to be useless [17].

Definition 3.0.2 (Useless classical queries [17]). *Let $(\mathcal{C}, \{\mathcal{C}_j : j \in J\}, \mu)$ be a learning problem. n classical queries are said to be useless, or to convey no information, if for any $x_1, \dots, x_n \in X$ and $y_1, \dots, y_n \in \{0,1\}$ the following holds*

$$\mu(f \in \mathcal{C}_j \mid f(x_i) = y_i, i = 1, \dots, n) = \mu(f \in \mathcal{C}_j), \text{ for all } j \in J.$$

A general n query algorithm in a generalised theory satisfying our principles corresponds to the following: an arbitrary initial state $|\sigma\rangle$ is prepared and input to the oracle \mathcal{O}_f , the

¹⁴One could alternatively consider replacing $\{0,1\}$ with a different finite set Y . This the result proved in this section, and in [17], also holds in the more general case.

¹⁵That is, if the probability that the chosen function belongs to a given subclass after n classical queries is the same as the known prior probability with which it was originally chosen.

¹⁶ $\lceil (N - 1)/2 \rceil + 1$ applications of the solution to Deutsch’s problem.

output state is acted upon by an arbitrary transformation G_1 independent of f , and the process is repeated. After the n th oracle query, the state is

$$|\rho_f\rangle = G_n \mathcal{O}_f G_{n-1} \cdots G_1 \mathcal{O}_f |\sigma\rangle.$$

The final step consists of measuring this state with an arbitrary measurement denoted as $\{(s|)\}_{s \in S}$. The final¹⁷ output of the algorithm is given by a map, which is independent of f from the set indexing the measurement outcome, S , to the set indexing the subclasses to which the function could belong, J . We can now generalise the definition of a useless quantum query from [17] to the case of generalised theories satisfying our principles.

Definition 3.0.3 (Useless generalised queries). *Let $(\mathcal{C}, \{\mathcal{C}_j : j \in J\}, \mu)$ be a learning problem. n generalised queries are said to be useless, or to convey no information, if for any n query generalised algorithm with initial state $|\sigma\rangle$, transformations G_n, \dots, G_1 , and measurement $\{(s|)\}_{s \in S}$ the following holds*

$$\mu(f \in \mathcal{C}_j \mid s) = \mu(f \in \mathcal{C}_j), \text{ for all possible } s \in S, j \in J.$$

We now present our main result, which generalises Theorem 1 from [17].

Theorem 3.0.4. *Let $(\mathcal{C}, \{\mathcal{C}_j : j \in J\}, \mu)$ be a learning problem. Suppose kn classical queries are useless. Then in any theory which satisfies our principles and has maximal order of interference k , n generalised queries are useless.*

We present the formal proof below, but first provide a rough sketch proof. In a theory with k th order interference, each state can be decomposed as in Eq. (1.1.1). Hence each state is explicitly indexed by subsets—of size at most $|I| = k$ —of the set of pure and perfectly distinguished states defining the oracle. Thus, after a single generalised query, the state is indexed by the valuation of the chosen function on at most k inputs. After n generalised queries, it is indexed by kn valuations. Therefore, a measurement can reveal at most kn valuations of the chosen function. But, as kn classical queries are useless, it must be that n generalised queries are also useless. The intuition behind this result is that, as a given state can have coherence between at most k basis states, one can use generalised superposition states to extract at most k valuations of a given function in a single query.

Proof. Our proof is essentially a slight generalisation of the original quantum one presented in [17]. We need to show that the probability of f being in \mathcal{C}_j does not change if outcome s is observed after n queries. That is, we must show

$$\sum_{f \in \mathcal{C}_j} \mu(f \mid s) = \mu(\mathcal{C}_j), \text{ for any } s \in S \text{ and } j \in J.$$

Using Bayes' theorem, we can write

$$\mu(f \mid s) = \frac{\mu(s \mid f) \mu(f)}{\mu(s)} = \frac{\mu(s \mid f) \mu(f)}{\sum_{g \in \mathcal{C}} \mu(g, s)} = \frac{\mu(s \mid f) \mu(f)}{\sum_{g \in \mathcal{C}} \mu(s \mid g) \mu(g)}$$

Then by noting that $\mu(s \mid h) = (s \mid \rho_h)$ and summing over f in \mathcal{C}_j , we have

$$\sum_{f \in \mathcal{C}_j} \mu(f \mid s) = \frac{(s \mid \sum_{f \in \mathcal{C}_j} \mu(f) \rho_f)}{(s \mid \sum_{g \in \mathcal{C}} \mu(g) \rho_g)}. \quad (3.0.1)$$

¹⁷As was noted by [17], the final transformation G_n is unnecessary, as it could be incorporated into the measurement.

Let's focus on $|\rho_f\rangle$. Given the decomposition in Eq. (1.1.1), every state can be written as

$$|\sigma\rangle = \sum_{I, |I|=1}^k \omega_I |\sigma\rangle = \sum_{I, |I|=1}^k \sigma_I.$$

Now, each $\mathcal{O}_f(\sigma_I)$ can depend on all $f(i)$ with $i \in I$. By padding out those I with $|I| < k$ with dummy indices, after a single query one can write

$$\mathcal{O}_f|\sigma\rangle = \sum_I \mathcal{O}_f(\sigma_I) = \sum_{T_1} Q_{T_1} \left(f(x_1^1), f(x_1^2), \dots, f(x_1^k) \right),$$

where the second equality is just a relabeling of the terms where $T_1 = \{x_1^1, x_1^2, \dots, x_1^k\}$ is the padded version of I , and hence each Q_{T_1} is a vector in the real vector space of states that depends on $f(x_1^1), f(x_1^2), \dots, f(x_1^k)$. Therefore, after n queries one can write the state as

$$|\rho_f\rangle = \sum_{T_n} Q_{T_n} \left(f(x_1^1), \dots, f(x_n^1), f(x_1^2), \dots, f(x_n^2), \dots, f(x_1^k), \dots, f(x_n^k) \right)$$

Using a change of variables provides

$$\begin{aligned} \sum_{f \in \mathcal{C}_j} \mu(f) |\rho_f\rangle &= \\ \sum_{T_n} \sum_{\{y_i^1\}, \dots, \{y_i^k\}} \mu(f \in \mathcal{C}_j \text{ and } f(x_i^m) = y_i^m, \text{ for } i = 1, \dots, n \text{ and } m = 1, \dots, k) & Q_{T_n} \left(y_1^1, \dots, y_n^k \right). \end{aligned}$$

As kn classical queries are useless

$$\begin{aligned} \mu(f \in \mathcal{C}_j \text{ and } f(x_i^m) = y_i^m, \text{ for } i = 1, \dots, n \text{ and } m = 1, \dots, k) &= \\ \mu(\mathcal{C}_j) \mu \left(f(x_i^j) = y_i^j, \text{ for } i = 1, \dots, n \text{ and } j = 1, \dots, k \right). \end{aligned}$$

Inputting this into the above we obtain,

$$\begin{aligned} \sum_{f \in \mathcal{C}_j} \mu(f) |\rho_f\rangle &= \\ \mu(\mathcal{C}_j) \sum_{T_n} \sum_{\{y_i^1\}, \dots, \{y_i^k\}} \mu \left(f(x_i^j) = y_i^j, \text{ for } i = 1, \dots, n \text{ and } j = 1, \dots, k \right) & Q_{T_n} \left(y_1^1, \dots, y_n^k \right). \end{aligned} \tag{3.0.2}$$

Then, summing over $j \in J$, results in

$$\begin{aligned} \sum_{f \in \mathcal{C}} \mu(f) |\rho_f\rangle &= \\ \sum_{T_n} \sum_{\{y_i^1\}, \dots, \{y_i^k\}} \mu \left(f(x_i^j) = y_i^j, \text{ for } i = 1, \dots, n \text{ and } j = 1, \dots, k \right) & Q_{T_n} \left(y_1^1, \dots, y_n^k \right). \end{aligned}$$

Substituting this back into Eq. (3.0.2) immediately gives

$$\sum_{f \in \mathcal{C}_j} \mu(f) |\rho_f\rangle = \mu(\mathcal{C}_j) \sum_{f \in \mathcal{C}} \mu(f) |\rho_f\rangle.$$

finally, substituting this into Eq. (3.0.1) completes the proof. \square

4 Conclusion

In this work we have introduced a well-defined oracle model for generalised probabilistic theories. This allowed us to compare the computational power imposed by different physical principles through the lens of query complexity. Our main result was to show that lower bounds on the number of queries to a quantum oracle needed to solve certain problems are not optimal in the space of generalised theories satisfying the principles introduced in section 1. Our result highlights the role of interference in computational advantages in a theory independent manner, demonstrating that “more interference implies more computational power”.

Previous work by two of the authors in [4] derived Grover’s lower bound to the search problem from simple physical principles. The search problem asks one to find a certain “marked item” from among a collection of items in an unordered database. The only access to the database is through an oracle; when asked if item i is the marked one, the oracle outputs “yes” or “no”. The figure of merit in this problem is how the minimum number of queries required to find the marked item scales with the size of the database. It was shown that, asymptotically, higher-order interference does not provide an advantage over quantum theory in this case. As opposed to the asymptotic behaviour of the number of queries needed to solve the search problem, the current work was concerned with whether a fixed number of queries yielded any information about the solution of particular query problem. In this case we were able to show that higher-order interference allows for a speed-up over quantum computation. Note that a specific oracle model for the search problem was introduced in [4]. However, this is just a special case of the general model introduced in section 2 of the current work. Moreover, the subroutine theorem proved here shows that our general oracle model is well-defined.

Our derivation of query lower bounds raises the question of whether the physical principles we have discussed are sufficient for the existence of algorithms which achieve these lower bounds. In the specific case of the search problem, a quantum search algorithm based on Hamiltonian simulation, presented in chapter 6 of the well known textbook by Nielsen and Chuang [15], may be more directly generalisable to theories satisfying our principles than Grover’s original construction [42]. This approach may also be applicable to many other query algorithms. In [15] they consider a Hamiltonian H consisting of projectors onto the marked item $|x\rangle$ and the initial input state $|\psi\rangle = \alpha|x\rangle + \beta|y\rangle$, with $|y\rangle$ orthogonal to $|x\rangle$ and $\alpha^2 + \beta^2 = 1$, respectively. That is, they consider the Hamiltonian $H = |x\rangle\langle x| + |\psi\rangle\langle\psi|$. Evolving the initial input state under this Hamiltonian for time t results in

$$\exp(-itH)|\psi\rangle = \cos(\alpha t)|\psi\rangle - i\sin(\alpha t)|x\rangle.$$

Hence, measuring the system in the $\{|x\rangle, |y\rangle\}$ basis at time $t = \pi/2\alpha$ yields outcome $|x\rangle$ with probability one. If the initial state was a uniform superposition over the orthonormal basis containing $|x\rangle$, then the required evolution time is $t = \pi\sqrt{N}/2$, where N is the size of the system (or equivalently, the number of elements in the database being searched).

One might wonder why there is no mention of an oracle in the above discussion. The oracle comes into play when constructing a quantum circuit to simulate the above Hamiltonian evolution. As the above Hamiltonian depends on the marked item, the quantum circuit simulating it must query the search oracle a number of times proportional to the evolution time [15]. In this specific case, an efficient Hamiltonian simulation requires $O(\sqrt{N})$ queries to the oracle, yielding an optimal quantum algorithm (up to constant factors) for the search problem. Recently, Barnum et al. [35] have introduced a physical principle, termed “energy observability”, which implies the existence of a continuous time evolution and ensures that the generator of such an evolution—a generalised “Hamiltonian”—is associated to an appropriate observable, which is a conserved quantity—the generalised “energy” of the evolving

system. Recall from section 1 that the principles we have discussed were sufficient to ensure that projectors onto arbitrary states correspond to allowed transformations. Hence, our previous principles, together with Barnum et al.’s energy observability, may be sufficient to run the above quantum search algorithm, hence providing a theory independent description of an optimal (up to constant factors) search algorithm. Similar constructions based on Hamiltonian simulation may also show that theories satisfying the above physical principles can reach the query lower bounds derived in this paper.

Acknowledgements

The authors thank D. Meyer for bringing to their attention his work on useless quantum queries with J. Pommersheim in [17]. The authors thank Matty Hoban for useful discussions and J.J. Barry for encouragement while writing the current paper. This work was supported by EPSRC grants through the Controlled Quantum Dynamics Centre for Doctoral Training, and the UCL Doctoral Prize Fellowship. We also acknowledge financial support from the European Research Council (ERC Grant Agreement no 337603), the Danish Council for Independent Research (Sapere Aude) and VILLUM FONDEN via the QMATH Centre of Excellence (Grant No. 10059). This work began while the authors were attending the “Formulating and Finding Higher-order Interference” workshop at the Perimeter Institute. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Research, Innovation and Science.

Appendices

A Proof of theorem 1.1.1

This is an adaptation of the proof of theorem 8 from [35]. Consider a self-dual cone \mathbf{C} with self-dualising inner product $\langle \cdot, \cdot \rangle$. Now consider a set of pure and perfectly distinguishable states ϕ_i which are distinguished by the effects e_i such that $(e_i|\phi_j) = \delta_{ij}$. We can define a face F of \mathbf{C} as the minimal face generated by the set of states $\{\phi_i\}$, we can moreover define the dual face $F^* := \{x \in \mathbf{C} \mid \langle x, s \rangle \geq 0 \ \forall s \in F\}$. Appendix A in [35] shows that if $F = F^*$ then there exists a positive projector onto the face F .

Consider some $t \in F$, self-duality of \mathbf{C} implies that $\langle t, s \rangle \geq 0 \ \forall s \in F$ hence, $t \in F^*$ and so $F \subseteq F^*$. We therefore just need to prove the converse inclusion and we are done.

To prove this, consider a normalised extremal $x \in F^*$, there must be some $s \in F$ such that $\langle s, x \rangle = 0$, where moreover, if $\langle s, y \rangle = 0$ then $y \propto x$. Next we prove two simple results:

i) s is not internal to F —assume, for the sake of contradiction, that s is internal. Then $\langle x, t \rangle = 0 \ \forall t \in F$ so $x = 0$ and hence is not normalised.

ii) There exists $t \in F$ such that s and t are perfectly distinguishable—assume, again to reach a contradiction, that there is no such t . This means that given any pure and perfectly distinguishing measurement $\{\epsilon_i\}$ that $(\epsilon_i|s) > 0$ for all i . Due to strong symmetry, any pure effect appears in such a measurement, therefore $(e|s) > 0$ for all pure effects $(e|)$, this suffices for tomography hence $|s\rangle$ is an internal state, in contradiction with (i).

Theorem 1 from [36] implies that if s and t are perfectly distinguishable states then $\langle s, t \rangle = 0$, therefore we know that $t \propto x$ and so $x \in F$. This is true for all extremal normalised $x \in F^*$ it therefore follows from convexity that this is true for all $x \in F^*$ and so we have $F^* \subseteq F$ which concludes the proof.

Hence, projectors P_F onto F are positive transformations. It was shown in [22] that in any theory satisfying causality, purification and computationally consistent composition, mathematically well-defined transformations are physical, i.e. they are allowed in the theory. Hence projectors P_F are physically allowed transformations. Moreover, given two faces, F and G , generated by different subsets of the same pure and perfectly distinguishable set of states, one has $P_F P_G = P_{F \cap G}$.

B Useful consequences of our principles

B.1 Uniqueness of distinguishing measurement

Strong symmetry (together with the no restriction hypothesis, which says that all mathematically well-defined measurements are physical) implies that, given any set of pure and perfectly distinguishable states $\{|i\rangle\}$, there exists a unique measurement $\{|j\rangle\}$ such that,

$$(i|j) = \delta_{ij}.$$

See [36, 35] for details. Moreover if there is a set $\{|e_j\rangle\}$ such that $(e_j|i) = \alpha_j \delta_{ij}$ then,

$$(e_j| = \alpha_j (j|.$$

B.2 Purification of the maximally mixed state is dynamically faithful

As mentioned in section 1, purification implies that there exists a *maximally mixed* state. Purification implies that there exists a state $|\psi\rangle$ that purifies the completely mixed state:

$$\left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \text{---} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) = \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \text{---} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right)$$

This is unique up to reversible transformation. We denote a particular choice of this purification as,

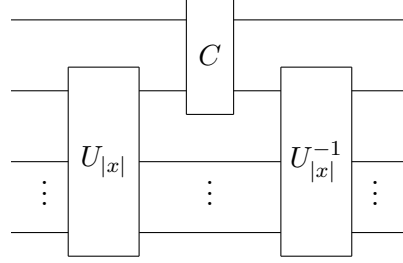
$$\left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) := \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \text{---} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right)$$

Purifications of the completely mixed state are called *dynamically faithful* states [22] and, due to the constraints on parallel composition imposed in section 1, must satisfy the following important condition [22]:

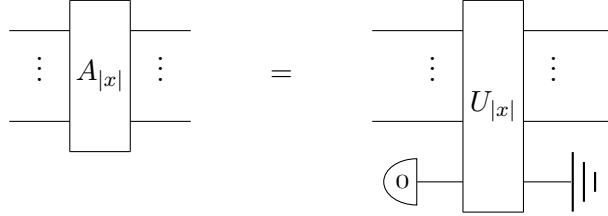
$$\begin{array}{c} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \text{---} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) = \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \text{---} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \\ \Rightarrow \\ \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \text{---} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) = \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \text{---} \left(\begin{array}{c} \text{---} \\ \text{---} \end{array} \right) \end{array}$$

C Proof of theorem 2.0.1

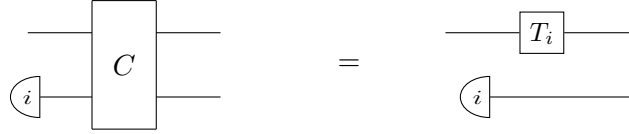
It was shown in [22] that the purification principle implies the ability to dilate any transformation to a reversible one. We use this fact in the construction of the circuit $\{G_{|x|}\}$. Our construction is equivalent to the one employed in the quantum case by [16]. Each $G_{|x|}$ corresponds to



where U is the reversible transformation which dilates¹⁸ the **BGP** algorithm $A_{|x|}$

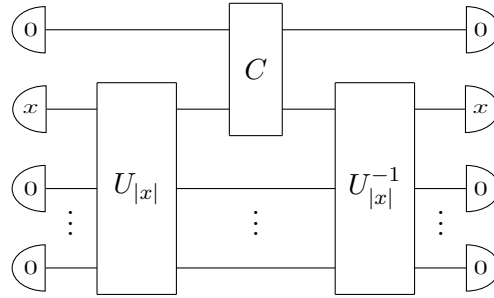


and C is a reversible controlled transformation with the lower system as the control



with $|i\rangle \in \{|0\rangle, |1\rangle\}$, $T_0 = \mathbb{I}$, and where T_1 acts as $T_1|i\rangle = |i \oplus 1\rangle$.

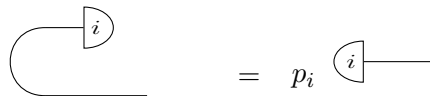
To prove theorem 2.0.1, we need to show that the probability corresponding to the following closed circuit



is greater than or equal to $1 - 2^{-q(|x|)}$, for some polynomial $q(|x|)$, when the algorithm accepts¹⁹ the input x , as this entails that $G_{|x|}$ functions as an oracle with high probability.

We now present the proof of theorem 2.0.1.

Proof. Choose the dynamically faithful state to satisfy



¹⁸Here we assume for simplicity that the circuit family $\{U_{|x|}\}$, with $U_{|x|}$ a reversible transformation which dilates $A_{|x|}$ for each $|x|$, consists of poly-size uniform circuits.

¹⁹That is, when x is in the required language decided by the algorithm.

where $p_i \in [0, 1]$ and $\sum_i p_i = 1$, which can always be achieved without loss of generality (see theorem 6 and corollary 9 from [22]). We first show that C satisfies

$$\text{Diagram of } C \text{ with two } 0 \text{ meters on inputs} = \text{Two } 0 \text{ meters in series}$$

Indeed, uniqueness of measurement (both of the following states give probability p_0 for $(0|(0|$, and probability zero for each of $(0|(1|$, $(1|(0|$, and $(1|(1|)$ implies

$$\text{Diagram of } C \text{ with } 0 \text{ meter on top input and feedback} = p_0 \text{ } 0 \text{ meter on top input}$$

From our choice of dynamically faithful state, it then follows that

$$\text{Diagram of } C \text{ with } 0 \text{ meter on top input and feedback} = \text{Two } 0 \text{ meters on top input and output}$$

Dynamical faithfulness then gives the required result. Write

$$\text{Diagram of } U_{|x|} \text{ with } x \text{ meter on top input and } 0 \text{ meters on other inputs} = \alpha \text{ } \sigma \text{ meter with vertical ellipsis dots}$$

where $|\sigma\rangle$ is a normalised state and $\alpha \in [0, 1]$. Our choice of acceptance condition, together with the fact that U is a dilation of the algorithm A , results in

$$\text{Diagram of } U_{|x|} \text{ with } x \text{ meter on top input and } 0 \text{ meters on other inputs} = \alpha = P_x(acc)$$

Combining these two results now gives

$$\text{Diagram of } C \text{ with } 0 \text{ meter on top input and } x \text{ meter on bottom input} = P_x(acc) \text{ } U_{|x|}^{-1} \text{ diagram} = P_x(acc)^2 \text{ } \sigma \text{ meter with vertical ellipsis dots}$$

where the last line follows from self-duality. Now, by amplifying the acceptance probability of the original algorithm A (see [2] for an in depth discussion of bounded error efficient computation and amplifying acceptance probabilities), we can ensure that when x is in the language we have $P_x(acc) \geq 1 - 2^{-p(|x|)}$ for an arbitrary polynomial $p(|x|)$. Hence it follows that $P_x(acc)^2 \geq 1 - 2^{-p(|x|)+1}$. If $(\sigma|\sigma) = 1$, choosing $p(|x|) = q(|x|) + 1$ completes the proof.

The case $(\sigma|\sigma) < 1$ can be easily accounted for. As $|\sigma\rangle$ and $\langle\sigma|$ can be efficiently prepared by a poly-size circuit, the factor $(\sigma|\sigma)$ can be approximated by a rational number to high accuracy (this is a consequence of the computational uniformity condition required to define computation in arbitrary physical theories, including quantum theory. See [2] and [3] for an expanded discussion of this point). Hence one can write $(\sigma|\sigma) = 1 - c2^{-w(|x|)}$, for w a polynomial in the size of the circuit and c a constant natural number. One can always find a polynomial q such that $(1 - c2^{-w(|x|)})(1 - 2^{-p(|x|)+1}) \geq 1 - 2^{-q(|x|)}$. This completes the proof. \square

References

- [1] R. Landauer, “Irreversibility and heat generation in the computing process,” *IBM journal of research and development*, vol. 5, no. 3, pp. 183–191, 1961.
- [2] C. M. Lee and J. Barrett, “Computation in generalised probabilistic theories,” *New Journal of Physics*, vol. 17, no. 8, p. 083001, 2015.
- [3] C. M. Lee and M. J. Hoban, “Bounds on the power of proofs and advice in general physical theories,” *Proc. R. Soc. A* 472 (2190), 20160076, 2016.
- [4] C. M. Lee and J. H. Selby, “Deriving grover’s lower bound from simple physical principles,” *New Journal of Physics*, vol. 18, no. 9, p. 093047, 2016.
- [5] J. Barrett, N. de Beaudrap, M. J. Hoban, and C. M. Lee, “The computational landscape of general physical theories,” arXiv:1702.08483, 2017.
- [6] C. M. Lee and M. J. Hoban, “The information content of systems in general physical theories,” *arXiv preprint arXiv:1606.06801*, 2016.
- [7] D. Stahlke, “Quantum interference as a resource for quantum speedup,” *Physical Review A*, vol. 90, no. 2, p. 022302, 2014.
- [8] C. M. Lee and J. H. Selby, “Generalised phase kick-back: the structure of computational algorithms from physical principles,” *New Journal of Physics*, vol. 18, no. 3, p. 033023, 2016.
- [9] R. D. Sorkin, “Quantum mechanics as quantum measure theory,” *Modern Physics Letters A*, vol. 9, no. 33, pp. 3119–3127, 1994.
- [10] R. D. Sorkin, “Quantum measure theory and its interpretation,” *arXiv preprint gr-qc/9507057*, 1995.
- [11] U. Sinha, C. Couteau, Z. Medendorp, I. Söllner, R. Laflamme, R. Sorkin, and G. Weihs, “Testing born’s rule in quantum mechanics with a triple slit experiment,” *arXiv preprint arXiv:0811.2068*, 2008.
- [12] D. K. Park, O. Moussa, and R. Laflamme, “Three path interference using nuclear magnetic resonance: a test of the consistency of born’s rule,” *New Journal of Physics*, vol. 14, no. 11, p. 113025, 2012.

- [13] U. Sinha, C. Couteau, T. Jennewein, R. Laflamme, and G. Weihs, “Ruling out multi-order interference in quantum mechanics,” *Science*, vol. 329, no. 5990, pp. 418–421, 2010.
- [14] T. Kauten, R. Keil, T. Kaufmann, B. Pressl, Č. Brukner, and G. Weihs, “Obtaining tight bounds on higher-order interferences with a 5-path interferometer,” *New J. Phys.* *19* 033017, 2017.
- [15] M. A. Nielsen and I. L. Chuang, *Quantum computation and quantum information*. Cambridge university press, 2010.
- [16] C. H. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, “Strengths and weaknesses of quantum computing,” *SIAM journal on Computing*, vol. 26, no. 5, pp. 1510–1523, 1997.
- [17] D. A. Meyer and J. Pommersheim, “On the uselessness of quantum queries,” *Theoretical Computer Science*, vol. 412, no. 51, pp. 7068–7074, 2011.
- [18] S. Aaronson, A. Bouland, J. Fitzsimons, and M. Lee, “The space just above bqp,” in *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*, pp. 271–280, ACM, 2016.
- [19] N. Bao, A. Bouland, and S. P. Jordan, “Grover search and the no-signaling principle,” *arXiv preprint arXiv:1511.00657*, 2015.
- [20] J. Barrett, “Information processing in generalized probabilistic theories,” *Physical Review A*, vol. 75, no. 3, p. 032304, 2007.
- [21] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, “Generalized no-broadcasting theorem,” *Physical review letters*, vol. 99, no. 24, p. 240501, 2007.
- [22] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Probabilistic theories with purification,” *Physical Review A*, vol. 81, no. 6, p. 062348, 2010.
- [23] L. Hardy, “Reformulating and reconstructing quantum theory,” *arXiv preprint arXiv:1104.2066*, 2011.
- [24] C. M. Lee and J. H. Selby, “A no-go theorem for theories that decohere to quantum mechanics,” *arXiv preprint arXiv:1701.07449*, 2017.
- [25] G. Chiribella and C. M. Scandolo, “Entanglement and thermodynamics in general probabilistic theories,” *New Journal of Physics*, vol. 17, no. 10, p. 103027, 2015.
- [26] B. Coecke and A. Kissinger, “Picturing quantum processes. a first course in quantum theory and diagrammatic reasoning,” *Cambridge University Press*, 2016.
- [27] B. Coecke, “Quantum pictorialism,” *Contemporary physics*, vol. 51, no. 1, pp. 59–83, 2010.
- [28] G. Chiribella and C. M. Scandolo, “Conservation of information and the foundations of quantum mechanics,” in *EPJ Web of Conferences*, vol. 95, p. 03003, EDP Sciences, 2015.
- [29] G. Chiribella and C. M. Scandolo, “Entanglement as an axiomatic foundation for statistical mechanics,” *arXiv preprint arXiv:1608.04459*, 2016.

- [30] G. Chiribella and C. M. Scandolo, “Purity in microcanonical thermodynamics: a tale of three resource theories,” *arXiv preprint arXiv:1608.04460*, 2016.
- [31] L. Disilvestro and D. Markham, “Quantum protocols within spekkens’ toy model,” *arXiv:1608.09012 [quant-ph]*, 2016.
- [32] R. W. Spekkens, “Evidence for the epistemic view of quantum states: A toy theory,” *Physical Review A*, vol. 75, no. 3, p. 032110, 2007.
- [33] G. M. D’Ariano, F. Manessi, P. Perinotti, and A. Tosini, “Fermionic computation is non-local tomographic and violates monogamy of entanglement,” *Europhys. Lett.*, vol. 107, no. 2, p. 20009, 2014.
- [34] G. M. D’Ariano, F. Manessi, P. Perinotti, and A. Tosini, “The Feynman problem and fermionic entanglement: Fermionic theory versus qubit theory,” *Int. J. Mod. Phys. A*, vol. 29, no. 17, p. 1430025, 2014.
- [35] H. Barnum, P. Markus, C. Ududec, *et al.*, “Higher-order interference and single-system postulates characterizing quantum theory,” *New Journal of Physics*, vol. 16, no. 12, p. 123029, 2014.
- [36] M. P. Müller and C. Ududec, “Structure of reversible computation determines the self-duality of quantum theory,” *Physical review letters*, vol. 108, no. 13, p. 130401, 2012.
- [37] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, “Quantum algorithms revisited,” in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 454, pp. 339–354, The Royal Society, 1998.
- [38] A. Sinha, A. H. Vijay, and U. Sinha, “On the superposition principle in interference experiments,” *Scientific reports*, vol. 5, 2015.
- [39] C. Ududec, H. Barnum, and J. Emerson, “Three slit experiments and the structure of quantum theory,” *Foundations of Physics*, vol. 41, no. 3, pp. 396–405, 2011.
- [40] C. M. Lee and J. H. Selby, “Higher-order interference in extensions of quantum theory,” *Foundations of Physics, Volume 47, Issue 1, pp 89–112*, 2017.
- [41] C. H. Papadimitriou, “Computational complexity,” *John Wiley and Sons Ltd.*, 2003.
- [42] L. K. Grover, “Quantum mechanics helps in searching for a needle in a haystack,” *Physical review letters*, vol. 79, no. 2, p. 325, 1997.